DOCUMENT RESUME

ED 370 984 TM 021 563

AUTHOR	Mehrens, William A.
TITLE	Issues and Recommendations Regarding Implementation of High School Graduation Tests.
INSTITUTION	North Central Regional Educational Lab., Oak Brook, IL.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	93
CONTRACT	RP91002007
NOTE	58p.; For a related document, see TM 021 564.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Administration; Educational Assessment; Educational Finance; Educational Planning; Educational Policy; *Graduation Requirements; *High School Graduates; High Schools; Human Resources; Legal Problems; *Psychometrics; *Test Construction; Test Reliability; Test Use; Test Validity
IDENTIFIERS	*Exit Examinations; High Stakes Tests; North Central Regional Educational Laboratory; United States (North Central)

ABSTRACT

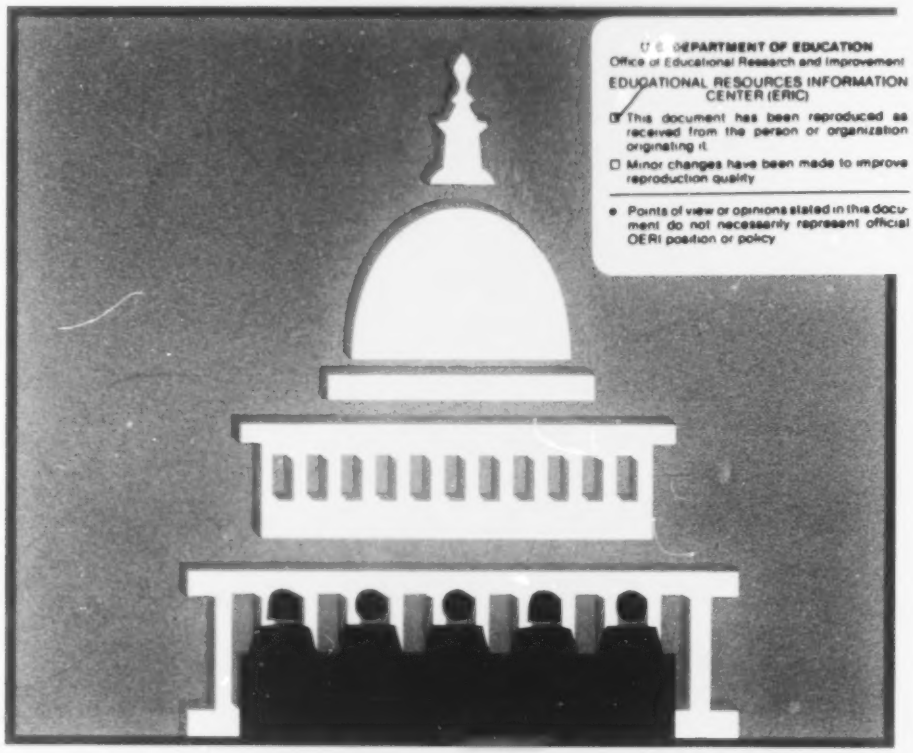
As the first paper in a series of policy papers on high-stakes student assessment programs, this paper examined high school graduation tests. High stakes refers to the use of test results to make important decisions about the test taker. Whether to use a high school graduation test is an essential policy question that will be addressed in a forthcoming paper; how to develop a sound graduation test is the focus of this paper, and most of the material is drawn from the experiences of the seven states participating in the North Central Educational Research Laboratory (NCREL). Brief descriptions of graduation requirements and tests are given. An expert panel employed by the NCREL to help Michigan develop its high school graduation test offers recommendations in the broad categories of (1) content specification, (2) psychometric issues, (3) educational issues, (4) legal issues, (5) policy and administrative issues, and (6) human and financial resources. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 370 984

Issues and Recommendations Regarding Implementation of High School Graduation Tests

REGIONAL POLICY INFORMATION CENTER


U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

by William A. Mehrens

TM021563

NCREL

BEST COPY AVAILABLE

NCREL

North Central Regional Educational Laboratory
1900 Spring Road, Suite 300
Oak Brook, IL 60521
(708) 571-4700, Fax (708) 571-4716

Jeri Nowakowski:	Executive Director
Deanna H. Durrett:	Director, RPIC
Lawrence B. Friedman:	Associate Director, RPIC
Linda Ann Bond:	Director of Assessment
John Blaser:	Editor
Stephanie L. Merrick:	Editor
Melissa Chapko:	Graphic Designer

NCREL is one of ten federally supported educational laboratories in the country. It works with education professionals in a seven-state region to support restructuring to promote learning for all students—especially those most at risk of academic failure in rural and urban schools.

The Regional Policy Information Center (RPIC) connects research and policy by providing federal, state, and local policymakers with research-based information on such topics as educational governance, teacher education, and student assessment policy.

© 1993 North Central Regional
Educational Laboratory

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number RP91002007. The content of this publication does not necessarily reflect the views of OERI, the Department of Education, or any other agency of the U.S. Government.

Table of Contents

Section I: **An Introduction** v
By Linda Ann Bond, NCREL's Regional Policy Information Center

Section II: **High School Graduation Requirements In the North Central Region** 1
Edited by E. Roger Trent, Ohio Department of Education

Section III: **Executive Summary** 9
By Linda Ann Bond

Section IV: **Issues and Recommendations Regarding Implementation of High School Graduation Tests** 17
By William A. Mehrens, Michigan State University

Foreword 19

Preface 21

Introduction 23

Issues and Recommendations 25

Core Curriculum/Test Specifications Issues 25
- Specify Subject Matter
- Specify Content Within Subjects
- Test a Sub-portion of the Core
- Determine Opportunity to Learn
- Phase-in Any Changes

Psychometric Issues 29
- Validity
- Item Development
- Field Testing
- Scoring
- Standard Setting
- Item Sensitivity Reviews and Bias Studies
- Reliability
- Scaling/Reporting
- Number of Forms
- Equating
- Standardization of Test Administration

Education Issues 39
Early Grade Testing
Retesting
Remediation
Special Education and Limited English Proficiency
Adult education

Legal Issues 41

Policy/Administrative Issues 44

Human and Financial Resource Issues 45
Staffing Needs
Advisory Committees
Contractors
Financial Resources

Sequence of Tasks 48

Conclusions 51

References 52

Appendix: Options 53

An Introduction

By Linda Ann Bond, Ph.D., Director of Assessment
NCREL Regional Policy Information Center

The Regional Policy Information Center (RPIC) of the North Central Regional Educational Laboratory (NCREL) offers the first in a series of policy papers concerning high stakes[1] student assessment programs—testing programs whose scores profoundly affect the lives of the students who take them and the lives of the educators and parents who want these students to be successful on them. Our intended audience is education policymakers and those who influence or are influenced by education policy decisions.

These papers offer a balanced presentation of the latest research-based and theory-based information. They do not provide "solutions"—those who work with education policy know that most policy decisions involve trade-offs. Instead, the papers describe the trade-offs in sufficient detail to assist policymakers in making informed decisions about high stakes student testing and assessment programs. For states that have already embarked upon high stakes testing programs, these papers describe the trade-offs inherent in different implementation strategies. As always, NCREL's primary interest is to offer information that will serve the best interests of learning for students, especially those students most at-risk of academic failure.

Because NCREL serves a seven-state region, priority will be given to the issues of greatest concern to policymakers in Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin. However, many of these topics are being considered by policymakers across the nation and are designed to be helpful beyond the NCREL region.

High School Graduation Testing

The first paper in the assessment series, *Issues and Recommendations Regarding Implementation of High School Graduation Tests*, examines high school graduation tests. In most states, individuals earn high school diplomas based on Carnegie units, which are defined as the number of hours the student has attended class. Instead of showing what students know at the end of high school, the current transcript simply reports which courses were taken and passed. Of course, the students receive grades in the courses they attend, but even in courses with the same title, the course content and the criteria used to define "A, B,

---

[1]High stakes refers to the use of test results to make important decisions about the test taker. For example, because high school competency tests can be used to deny students a property right—a diploma—these tests fall into the "high stakes" category. Due to the importance of high stakes tests, they must be of the highest quality and must be able to withstand any challenge in court.

C, D, or F" can vary considerably. For this reason, states are considering setting uniform performance expectations that all high school students must meet and are developing test or assessment systems to certify satisfactory performance.

This first paper on graduation tests has been revised from one commissioned by the Michigan Department of Education, funded by NCREL, and written this spring by a panel of experts (listed in the complete paper that follows) who were called in to help guide the implementation of Michigan's newly mandated high school graduation test. Their paper, though intended for the state board of education in Michigan, deals with the broader issues that any state needs to address when considering a high school graduation test. By revising the paper slightly to maximize its utility for all states in the NCREL region and beyond, we believe the paper merits the attention of policymakers involved in decisions concerning graduation testing or any other assessment approach intended to eliminate the high school's reliance on the Carnegie unit to measure student success in high school.

Whether to use a graduation test is an essential policy question, and its ramifications will be addressed in a forthcoming volume in this series. How to develop a sound graduation test once the decision has been made to do so, is also a key policy question—the one that this paper addresses. Because a majority of the seven states in the NCREL region have already implemented a graduation test (Ohio), are moving toward such a test (Indiana and Michigan), or are considering the possibility (Wisconsin), the latter question is addressed first. Included in this paper is a description of graduation requirements in the NCREL region. This section is followed by an executive summary of the regionalized Michigan paper, followed by the paper itself. Future papers will deal with the pros and cons of high school graduation testing and the legal implications of high stakes testing.

High School Graduation Requirements In the North Central Region

Edited by E. Roger Trent, Ohio Department of Education

The NCREL region boasts a long tradition of local control. Several states in the region allow local school districts to determine completely the graduation requirements for their students (Illinois and Iowa), while the others set at least some graduation requirements at the state level. While most of the states in the region still rely upon Carnegie units (courses taken and passed) as the measure of successful completion of high school, several are exploring strategies to move away from this "inputs-based" system into one that is outcomes-based. High school competency testing is being used as one approach to accomplish this goal. Ohio has developed such a test already, Michigan and Indiana both have legislative mandates to develop the same type of test, and Wisconsin is developing a tenth-grade "gateway exam" that may someday be used as a high school graduation test.

The following section provides, for each state in the region, brief descriptions of graduation requirements, including tests and any initiatives to award high school diplomas based on demonstrated competencies, rather than Carnegie units.

Illinois

Illinois maintains complete local control of high school graduation requirements. Each school district determines its own requirements and issues its own high school diplomas. The state does not set requirements for graduation, such as passage of proficiency tests.

The state has no plans to alter its system of local control. No bills calling for graduation tests are pending in Illinois. Nor is the state expected to enact legislation or administrative rules that would require either the state or local districts to have students demonstrate certain competencies to earn a high school diploma.

However, Illinois does not take an entirely hands-off approach to setting high school graduation requirements. Although educational accountability remains at the school level, schools are required to meet the competencies defined in the Illinois Goal Assessment Program (IGAP) series of tests. Schools also must follow guidelines based on the State Goals for Learning as they establish and measure outcomes through their local assessment systems.

Contact: Carmen Chapman, Supervisor of Professional Development
Illinois State Board of Education
217-782-4823

Indiana

Current Graduation Requirements Although the governing board of each school corporation in Indiana issues high school diplomas, the state sets requirements for high school graduation. The state requirements include 38 Carnegie unit credits—eight units in language arts; four units each in mathematics, science, and social studies; and one unit each in physical education and health and safety. Students also must attend at least seven semesters in grades nine though 12, unless the requirement is waived in accordance with specific criteria of the Indiana Board of Education. Performance standards do not exist, but local school boards, with the approval of the state board of education, may set such standards to reflect "competency in the basic skills necessary for future learning."

A second type of certificate—an Academic Honors diploma—also is available to Indiana students who earn at least 47 credits and take more rigorous coursework in specific academic subjects. For example, students eligible for this diploma may be required to take advanced math courses such as Algebra II.

Future Graduation Requirements During the 1992 Indiana legislative session, a Work Force Development Bill (Senate Bill 419) that will profoundly affect the awarding of high school diplomas was signed into law. The new law, which will take effect during the 1994-95 academic year, includes the following provisions:

> Grade 10 Gateway Exam and Gateway Certificates: All tenth-grade students will take a gateway exam that will yield both individual and school-based scores. A State Standards Task Force—comprising representatives from education, business, and labor—will recommend standards and content for the exam to the state board of education. Students will be expected to pass the exam and receive a Gateway Certificate as one requirement for graduation, although exceptions will exist for special education students and students in need of an alternative form of assessment. Remediation will be provided if state funds permit.
>
> Grades 11 and 12 Options for Students: Students who pass the gateway exam will be expected to develop a career plan and choose a technical or college preparatory curriculum for the remainder of their high school careers. This career plan will be developed in cooperation with a guidance counselor and the student's parents.
>
> Technical Certificates of Achievement: A student who chooses the technical preparatory curriculum will be required to pass a state-selected technical assessment and receive a Technical Certificate of Achievement in his/her field of study. These certificates can be made a graduation requirement at the discretion of the individual school corporation's governing board.

Academic Certificates of Achievement: A students who chooses the college preparatory curriculum may take Advanced Placement exams in a variety of courses and receive and Academic Certificate of Achievement.

Alternative Education: School corporations may develop an alternative program for students who fail to obtain the Gateway Certificate. The alternative program must be approved by the state board of education.

The state board of education is considering strategies for implementing this new law. The State Standards Task Force must submit its recommendations for the gateway exam to the Board by January 1993.

Contact: Richard Peters, Director of Student Assessment
Indiana Department of Education
317-232-9050

Iowa

Graduation requirements are established and diplomas issued by the board or governing authority of each school district. Boards that provide an education program through grade 12 must adopt a policy specifying graduation requirements, including provisions for early graduation.

The state does not require—nor is legislation pending that would require—students to pass proficiency tests to graduate from high school. However, the 1992 Iowa General Assembly has appointed an interim study committee to recommend goals and necessary reform legislation, including suggestions for alternative approaches to student assessment. The final report is due December 1, 1992.

In the meantime, the Iowa State Department of Education has undertaken a consensus building process to identify a limited number of broad exit outcomes. Although the initiative constitutes an initial step toward defining statewide outcomes, these outcomes are not expected to be tied specifically to graduation requirements.

Under the initiative, local districts would select assessment instruments related to local goals and outcomes. In turn, the state would develop an indicator system for monitoring attainment of statewide outcomes. Although the state does not intend to award diplomas based on the attainment of student outcomes, the local districts have discretion to do so.

Contact: Leland Tack
Division of Financial and Information Services
Iowa Department of Education
515-281-5293

Michigan

Civics is the only course that state law require students to take. A student who has not successfully completed this course will not be issued a diploma, unless the student has enlisted or has been inducted into military service. All other high school graduation requirements are established and diplomas issued by the local board of education of each school district.

In March 1990, the Michigan State Legislature enacted a set of broad educational reform measures for Michigan education. Section 1278 of Public Act 25 states that a recommended model core curriculum shall be developed by the state board of education and distributed to each school district in the state. The recommended core curriculum is based on the "Michigan K-12 Program Standards of Quality" and defines outcomes to be achieved by all students. Local school districts are asked to use the model core curriculum outcomes as a guide in developing their own core curriculum outcomes. The model core curriculum shifts the emphasis from *what is taught* to students to *what is learned* by students. It also creates a new accountability for education based on results, not intentions.

In the fall of 1991, as part of the State School Aid Act, the Michigan Legislature enacted two high school graduation requirements. The Act requires all public schools to award state endorsements on the high school diplomas of pupils who meet certain criteria. These pupils must graduate in 1994, 1995, or 1996 and must achieve a certain score on the Michigan Educational Assessment Program (MEAP) tests in mathematics, reading, and science or a locally adopted, state-approved test in these content areas. The MEAP mathematics and reading tests are administered in the fall of the tenth-grade school year, and the MEAP science test is given in the fall of the eleventh grade. Pupils have multiple opportunities to take the test during their high school years. They also have the option of passing the GED.

The State School Aid Act also requires schools to offer proficiency testing as a prerequisite for high school graduation:

> Not later than July 31, 1993, the department shall develop and the state board shall approve assessment instruments to determine pupil proficiency in communications skills, mathematics, science and other subject areas specified by the state board. The assessment instruments shall be based on the state board model core curriculum outcomes. Beginning with the graduating class of 1997, a pupil shall not receive a high school diploma unless the pupil achieves a passing score on the assessment instruments developed under this section. (Section 104(a)7)

The assessment instruments will be administered for the first time to the class of 1997 in the spring of 1995.

Contact: Peggy Dutcher
Michigan Department of Education
517-373-8393

Minnesota

In order to graduate, students must successfully complete 15 credits in a three-year secondary school or 20 credits in a four-year secondary school. Satisfactory completion of at least 120 hours is the basis for a credit course. Districts must provide students with the opportunity to earn at least six credits per year in grades 10, 11, and 12. The following common branches of learning—or subjects—and credits are required: four credits in communication skills; three in social studies; one each in mathematics and science; one-half in health; two-thirds in physical education at grade nine; one-half in physical education at grade 10; and the remaining credits selected by the student.

Over the past two years, the Minnesota State Board of Education has been developing an outcomes-based graduation rule to replace the current rule based on Carnegie units. During the 1992 legislative session, the legislature declared its commitment to establishing a rigorous, results-oriented graduation rule for the state's public school students according to the following time line:

- The state board of education will report to the legislature by February 1, 1993, and January 1, 1994, on the proceedings to adopt a graduation rule.

- Final action to adopt the rule may not be taken until July 1, 1994.

The legislature precludes the state board of education from prescribing the delivery system, form of instruction, or a single statewide form of assessment that local sites must use to meet the requirements contained in the rule.

The proposed rule, in its current form, establishes outcomes for the graduates of the year 2000 in three areas: reading, writing, and mathematical processes. Additional outcomes will be added in the year 2001. The Minnesota Department of Education will develop models of assessment that the district may adopt to verify that the outcomes have been attained. Districts will be responsible for verifying student learning and their plan/program will be subject to department review and approval.

Contact: Joan Wallin
Minnesota Department of Education
612-296-1570

Ohio

High school diplomas are issued by each of Ohio's 612 school districts. According to state high school graduation requirements, students must earn a minimum of 18 Carnegie units—three in language arts, two each in mathematics and social studies (including one-half each in United States history and government), one-half unit each in health and physical education, nine elective units, and a minimum of three units in a subject other than language arts. Many school districts have established higher requirements. None can have lower requirements.

To graduate after September 15, 1993, each student also must pass state proficiency tests in reading, writing, mathematics, and citizenship. Science proficiency tests will be developed and administered for the first time to ninth graders in the 1995-96 school year. Beginning in the 1998-99 school year, students who receive a diploma must pass a science test along with tests in the other four subject areas.

Students first take the state proficiency tests in November of their ninth-grade year. They may retake any failed test(s) in March. Makeup exams are offered twice each year until students have passed all required tests.

Students with identified disabilities must take the tests unless exemptions are granted through their individual education plan (IEP). Modifications in test administration procedures and format (e.g., large print edition) may be made in accordance with the IEP. Students with limited English proficiency, with parental permission, may not be required to take tests until acquiring minimal language proficiency. However, these students are not exempted from the requirement to pass the tests to earn a diploma.

Outcomes measured by the tests were adopted by the state board of education in 1988, after a year-long consensus-building process involving thousands of Ohio citizens. Performance standards for the first form of the tests were recommended by groups of ninth-grade teachers, reviewed by various panels of educators, and adopted by the state board of education. Passing standards for each succeeding test form are equated statistically to those established by the board.

Individual student results are used solely as part of the criteria for graduation. Aggregate percentages of students passing each test area are reported publicly after each test administration. These passing rates are used to identify excellent and deficient schools and districts.

In addition, the state board of education has announced its intent to initiate a system of awarding diplomas based on demonstrated competencies rather than Carnegie units. The board is studying a variety of implementation models. Although any recommendation is likely to include reference to the high school competency test, the relationship between this test and the new system is yet to be determined.

Contact: E. Roger Trent
Division of Educational Services
Ohio Department of Education
614-466-3224

Wisconsin

Wisconsin diplomas are issued by the 382 local school districts. However, the state has established requirements for high school graduation. They include the following:

- Four credits of English, including writing composition, oral communication, grammar, and usage of English language and literature
- Three credits of social studies, including state and local government
- Two credits each of mathematics, which incorporate instruction in the properties, processes, and symbols of arithmetic and elements of algebra and statistics, and science, which incorporate instruction in the biological and physical sciences
- One and one-half credits of physical education
- One-half credit of health education (in grades 7-12)

School boards are encouraged to require an additional eight and one-half credits of course work.

Wisconsin does not require students to pass state-level proficiency tests as a condition of high school graduation. Although no pending legislation calls for the use of state tests as a graduation requirement, this issue may arise after more work is completed on the "Gateway Assessment" program. Students who pass this high school proficiency exam elect coursework in either a college-preparatory or technical area.

During 1992-93, tenth-grade learner outcomes will be developed, and these may be considered part of the graduation requirements when a new assessment program becomes fully operational in 1996-97.

The state is just beginning the process of establishing state education goals, identifying learner outcomes, and designing a new performance-based assessment system.

Contact: Darwin Kaufman
Bureau for Student Assessment
Wisconsin Department of Public Instruction
608-266-9111

Issues and Recommendations Regarding Implementation of High School Graduation Tests

An Executive Summary

By Linda Ann Bond, Ph.D., Director of Assessment
NCREL Regional Policy Information Center

Introduction

High school graduation tests, or minimum competency tests, became popular in the United States during the late 1970s and mid-1980s as a response to the warning most clearly expressed in *A Nation at Risk* (1983)—a warning that schools were not providing students with the skills needed for success. The high school graduation test was seen as a means to ensure that high school graduates possessed a satisfactory level of basic skills (most often, reading and mathematics) needed for success in the community and the workplace.

A new wave of educational reform in the 1990s has brought with it a resurgence of interest in high school graduation tests, but the types of skills that are now deemed essential to success have changed. Instead of holding students to "minimal" skills, these new mandates are intended to raise standards beyond minimal levels of achievement. Current thinking suggests that to be successful in today's technologically advanced workplace, high school graduates need skills that used to be reserved for the college-bound. Minimum competencies are not enough. Many policymakers today look to graduation tests to raise the high school graduate's skills and knowledge to the higher level expected for success in a complex, demanding society and workplace.

Because a high school graduation test carries with it such high stakes, careful attention to the soundness of the test design process and to the legal defensibility of the test product is of critical importance. In this paper, the expert panel employed by NCREL for the Michigan State Board of Education offers its recommendations based on agreed upon testing standards. Their recommendations grow out of a wealth of experience with high stakes testing.

What are their recommendations? They advise states to move slowly and to document every step of the design and implementation process. They caution against attempting to develop the test without sufficient staff and resources and recommend some benchmark definitions of "sufficient." Technical standards (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1985) must be applied to the assessment(s), and the tests must be fair and consistently applied to all who take them. These standards are not as well-defined for the newer, nontraditional assessments, and the panel advises against using these assessments—especially when they involve observing and rating performance levels—for high stakes purposes until and unless sufficient research has demonstrated their effectiveness for this purpose.

Legally, students and their parents must be informed about the test requirement and the content of the test by the time the student is in the ninth grade. Students must receive instruction in the knowledge and skills included on the test prior to its implementation. Unless clearly addressed in the law, the state board of education should adopt specific test-taking procedures for special education and limited English proficient students, including exemption, special administrative adaptations, and adapted versions of the test.

This paper does not comment on the appropriateness of a high school graduation test as education policy. Instead, it focuses on the practical policy issues that a state should consider as it moves toward adopting a high school graduation test. These issues are divided into five broad categories: curriculum/content specification; psychometric issues; education issues; legal issues; and human and financial resources.

I Content Specification

Selecting the knowledge and skills that should be included on the test is one of the most important and most debated decisions to be made about any high stakes test. The curriculum should serve as the guide, and the test should include only content that students have had the opportunity to learn.

A. Start Small Many fear that subjects not included on the graduation test will not be taught. However, with tight legislative timelines and limited resources, it is best to start with only those subjects specified in the law or, if not specified, those considered most important.

B. Sample the Content Domain Again, many have expressed legitimate concern that what is not on the test will not be taught. Teaching to the test must be minimized by ensuring that test security measures are taken and that the sample covered by the assessment is *not* merely a minimal set of objectives or a "lowest common denominator" of what schools already teach. It is feasible and defensible to allow the test to lead the schools to some extent, and test developers must be careful to choose content that can be measured adequately and that students have had an opportunity to learn.

C. Provide the Opportunity to Learn One of the primary reasons to specify a core curriculum for coverage on the state graduation test is the recognition that, to be fair, every student required to take the test must receive instruction in all of the skills assessed by the test. "When a test is used to make decisions about student promotion or graduation, there should be evidence that the test covers only the specific or generalized knowledge, skills, and abilities that students have had the opportunity to learn" (AERA, APA, NCME, 1985, p. 53).

D. Obtain Evidence and Provide Support to Ensure That Every Student Has the Opportunity to Learn Gather evidence from teachers and students to indicate that

what is being tested is indeed being taught. States should provide teachers with professional development opportunities as needed.

E. Phase-in of Any Changes The graduation testing program is bound to change over time, but should do so gradually to protect students' opportunity to learn and educators' readiness to teach. Any changes should be considered in light of items A through D above.

II Psychometric Issues

Test construction, administration, and scoring and reporting of results should be governed by the *Standards for Educational Psychological Testing* (AERA, APA, NCME, 1985). This section of the paper deals with issues such as validity, item development, field testing, standard setting, item sensitivity reviews and bias studies, reliability, scaling/reporting, number of forms, equating, and standardization of administration.

A. Validity Validity is critical to any test used for graduation purposes. It requires evidence that the test measures what it purports to measure and that the inferences made from the test scores are justified. Although the paper goes into some detail about different types of validity, two major issues must be addressed.

First, does the test sample the content being tested sufficiently to justify its name—e.g., reading test, writing test, literacy test? The test developer must ensure that the test adequately samples the defined content.

Second, does evidence support claims about the test results? If the test claims to certify that an individual will be successful once he/she leaves school, research evidence must show that high scorers perform better in post-high school life than low scorers. An official statement from the department of education and/or the state board of education should caution against making unjustified inferences from the test scores.

B. Item Development Test developers must write a sufficient number of high quality test items to allow for losses due to pilot testing and to build enough test forms to sustain the program through its first two years. Item format (i.e., multiple-choice, essay, performance tasks) must be considered in light of content coverage, resources, research evidence of technical quality, and pilot testing.

C. Field Testing If at all possible, administer a field test in the first year, unless this would seriously jeopardize test security. A field test with a sample of students is essential prior to using the test for high stakes purposes.

D. Scoring Scoring must be accurate because of the importance of the results. Therefore, scoring methods must guarantee accuracy. (See Reliability section.)

E. Standard Setting Since this issue is one that frequently brings states to court, the standards must be set in a legally defensible fashion. It is important to ensure that the population of students who fail the test does not include large numbers of students who teachers believe should have passed. It is strongly recommended that a trained standard-setting committee be appointed along with a technical advisory committee to conduct the standard-setting procedure and document its appropriateness. Standards should be based on the first live administration of the test rather than the pilot, since students tend to take the former more seriously. If a high cut score is chosen, states might consider setting incremental cut scores for different graduating classes.

F. Item Sensitivity Reviews and Item Bias Studies All items to be used on a graduation test should be free of ethnic, cultural, and gender bias. A committee of individuals, including representatives of various groups, must be trained to review the test items to ensure that language and content do not favor a particular group. At least one member of the committee should be a minority group member from another state who is a recognized expert in this area. Item bias studies also should be conducted to demonstrate that items do not function differently in different subgroups. Any items that do function differently should be brought back to the item review committee for re-examination. It is not necessary that all subgroups have the same average scores to declare the test bias-free.

G. Reliability Test scores should reflect differences in the knowledge and skills of the test-takers, not irrelevant factors such as scoring errors and test-item familiarity. Reliability estimates are needed for internal consistency (e.g., high scorers perform equally well on all parts of the test, as do low scorers), inter-rater reliability (two trained scorers come up with approximately the same score for the same individual), generalizability across writing samples (doing well on one writing sample means doing well on others), and the reliability or standard error at the cut score (students who are above or below the cut score are accurately placed).

H. Scaling/Reporting Scores should be reported as "pass" or "fail." Those individuals who "fail" should be given some information regarding how close they were to passing and should be given some useful information for remediation purposes. The scale should be determined by a technical advisory panel. It may be helpful for interpretation purposes to use the same scale for all subject matter areas.

I. Number of Forms/Equating Sufficient test forms should be available to avoid using the same items each year. However, the difficulty of each form of the test and the content covered must be comparable from year to year to ensure fairness.

J. Standardization of Administration To be sure that all students have the same administration procedures, local school personnel must be trained to administer the tests, and random auditing should be conducted to ensure uniformity throughout the state.

III Educational Issues

Of obvious concern to any state implementing a high school graduation test is the assistance that should be offered to students who fail to achieve a satisfactory score on the test. Following are several recommendations for effectively addressing this concern.

A. Early Grade Testing If a graduation test is being considered, a state might also consider having earlier grade testing to identify and help students who may not be acquiring prerequisite knowledge and skills at the expected rate. Do not promise students that these scores will predict high school performance unless evidence suggests that they do.

B. Retesting Specific state board of education rules should govern retesting opportunities across the state. Students who are unable to pass the test after four or five attempts should be given the unlimited opportunity to retake the exam through an adult education program.

C. Remediation A state that adopts a diploma sanction test requirement should be responsible for assisting the local schools in planning for remediation. The respective responsibilities of the state, the district, and the student for remediation efforts should be clearly delineated.

D. Special Education and Limited English Proficiency Unless clearly addressed in the law, the state board of education should adopt specific test-taking procedures for special education students, including exemption, special administrative adaptations, and adapted versions of the test. Similarly, test-taking procedures for limited English proficient students should be specified, including whether the test should be administered in the student's first language. The attorney general should be consulted when making these decisions.

E. Adult Education Students in adult education programs who want to receive a high school diploma should be given the opportunity to take the high school graduation test.

IV Legal Issues

In general, it is wise to involve the state attorney general early in the test development process. Many of the legal issues concerning high school graduation testing were addressed in the case of *Debra P. v. Turlington* (1983, 1984), a broad-based challenge to Florida's graduation test.

A. Technical Soundness The standards (AERA, APA, NCME, 1985) mentioned in the psychometric section of this paper determine the technical soundness of a

graduation test if it is challenged in court. It is important to have documentation of the process used to select test content, to prevent bias, to assure reliability and validity, to set the cut score, and to ensure standardization of administration.

B. Liability It is important to determine the liability of the staff and advisory committees prior to implementation. Are teachers who help with the test or who are sued for not teaching the curriculum covered on the test protected from liability? Those involved in the test design and implementation process should be advised of their liability.

C. Due Process Individual students need sufficient notice of new graduation requirements, including information about the content of the test. Parents also must be notified and documentation of notification should be kept. Students and parents should be notified when the students are in ninth grade or even earlier.

D. Complete Records All steps in the design and implementation process should be kept on file for at least five years. Documentation of initial development should be kept indefinitely. Detailed policies regarding what should be documented and for how long must be determined and understood by everyone working with the test.

V Policy/Administrative Issues

States will need to establish a uniform set of rules for test administration. It is important to document the formal procedures used to establish these rules and to inform schools about the rules. Many challenges to test administration result from inadequate documentation of the rule-making procedures.

A. Administrative Rules One of the best ways to document that appropriate steps have been taken in developing a high school graduation test is to promulgate state board of education administrative rules. These rules should deal with issues such as frequency and timing of the test, rescoring policies, procedures for handling transfer students and special students, issuance of RFPs, and security. Test security is especially important. Secure tests should be excluded from freedom of information laws, but, if they are not, new legislation should be sought.

B. Frequency of Administration An annual test schedule should be developed and disseminated to all school districts. A graduation test should be given first in the spring of the tenth grade school year, and twice more in the junior and senior years.

VI Human and Financial Resources

Any test that has as profound an impact on students as does a high school graduation test must be well-conceived and carefully implemented. Sufficient staff and financial resources are necessary to satisfy the many technical and legal requirements of these kinds of tests.

A. Staffing Needs Because a graduation test has such serious implications for students, the time required for development and implementation is substantial. As a rule of thumb, the expert committee recommends one staff person for each major content area to address test content issues, one measurement specialist who can write RFPs to address psychometric issues, one individual whose sole task is to manage the contract and monitor the contractor, one person to manage staff, and an overall supervisor to ensure that all of the important work is done.

B. Advisory Committees A state needs to have input from local educators and content/technical experts throughout the test design and implementation process. This panel recommends the following: (1) a department of education steering committee, (2) a testing policy advisory committee, (3) a bias review committee, (4) a technical advisory committee, (5) a content review committee for each content area, (6) an overall content review committee, and (7) a standard setting committee.

C. Contractors Limit the number of separate contractors to two—one to work on test development, the other for test administration, scoring, and reporting.

D. Financial Resources Sufficient resources in staff and money are needed to do the high quality job required to defend the test legally. The paper offers a benchmark figure for test development over a two-year period ($650,000), but this figure presumes that the number of staff recommended above is already funded for the project. Financial and staffing information should be collected from states that have developed legally defensible graduation tests to support a state education agency's request from the legislature.

VII Summary

This paper offers the advice and recommendations of a group of testing and legal experts who were brought together to advise Michigan about the development of a high school graduation test. Although the content of the paper is specific to Michigan, the recommendations are relevant to any state that is beginning to develop a high school graduation test.

Issues and Recommendations Regarding Implementation of High School Graduation Tests

Spring 1992

A Report Prepared for the
North Central Regional Educational Laboratory
by an Expert Panel on the Michigan High School Graduation Test

Written by
William A. Mehrens, Michigan State University

NOTE: The contents of this report have been adapted, with permission, from a report written for the Michigan Department of Education by an Expert Panel whose members are listed below. Neither the contents of this report nor the original report necessarily represent the official opinions of any of the agencies that employ the members of this Expert Panel.

Expert Panel Members

Thomas Fisher, Administrator
Student Assessment Services
Florida Department of Education
FEC Building, Suite 701
325 Gaines Street
Tallahassee, FL 32399

Sharon Johnson-Lewis, Director
Planning, Research, and Evaluation
Detroit Public Schools
5057 Woodward
Detroit, MI 48202

Marjorie Mastie
Supervisor for Assessment Services
Washtenaw I.S.D.
1819 South Wagner Road
Ann Arbor, MI 48103

William Mehrens, Expert Panel Chair
Professor of Educational Measurement
462 Erickson Hall
Michigan State University
East Lansing, MI 48824-1034

Jason Millman
Professor of Educational Measurement
405 Kennedy Hall
Cornell University
Ithaca, NY 14853-4203

S.E. Phillips
Associate Professor of Education
458 Erickson Hall
Michigan State University
East Lansing, MI 48824-1034

Edward Roeber
Director of Student Assessment Programs
Council of Chief State School Officers
One Massachusetts Ave, N.W., Suite 700
Washington, DC 20001-1431

E. Roger Trent, Director
Division of Educational Services
Ohio Department of Education
65 South Front Street, Room 811
Columbus, OH 43266-0308

Foreword

The Expert Panel on the Michigan High School Graduation Test was convened by Interim Superintendent Gary D. Hawks to advise the Michigan Board of Education on important issues surrounding the high school proficiency examination enacted by Public Act 118 of 1991, Section 104a (Subsection 7). The panel members are national experts who have first-hand knowledge and experience with large-scale competency testing programs; they brought to the meetings a wealth of information and wisdom on the challenging issues that Michigan will face as it implements the provisions of the Act.

The Expert Panel met over three days in February and March of 1992 to examine the educational, technical, legal, fiscal, and logistical issues relating to competency testing. This report lists and provides the rationale for 51 recommendations to the state board of education for a technically and legally sound high school proficiency examination program within the time limitations that the legislation provides.

The Michigan Department of Education appreciates the assistance provided by the Expert Panel members. The work of this panel was made possible by a grant from the North Central Regional Educational Laboratory. Their support is gratefully acknowledged.

Preface

The purpose of this report is to provide advice to the state departments of education in the North Central Region regarding the issues and recommendations to be considered in developing and implementing a high school graduation test. In this report, the author discusses issues that need to be resolved, offers recommendations, and presents an illustrative list of tasks to be performed, with suggested completion dates.

The report assumes that a legislative act calls for the department of education to develop and the state board of education to approve assessment instruments to determine pupil proficiency. These assessment instruments are to be based on a core curriculum and a pupil shall not receive a high school diploma unless he/she achieves passing scores on these instruments.

Certainly it is possible to develop a high school graduation test that meets curricular, psychometric, educational, legal, administrative, and resource requirements. However, as this document makes clear, the task is not easy and time-lines are frequently tight. For the task to be done well, a variety of steps need to be taken soon after any legislative enactment. Immediate funding will be needed to ensure adequate human and fiscal resources. Only with appropriate funding to complete the task will a high school test graduation requirement be of service to the citizens of a state.

Introduction

The purpose of this report is to offer advice on the issues that need to be considered (and resolved) and the steps that need to be taken when implementing high school graduation tests. The paper also discusses advantages and disadvantages of potential decisions.

It is believed that a general report will be most useful if readers can see the advice and recommendations that stem from a *specific context*. Thus, the major portion of this report has been adapted, almost intact, from a report written for the Michigan Department of Education. That report was the product of an eight-member expert panel's deliberations (chaired by the author of this more general report) and was a response to a request for advice on how to implement a specific public act requiring a high school graduation test.

Obviously, the advice was given within a specific context created by certain variables mentioned in the legislation. Some of the more relevant contextual factors can be summarized as follows:

- The legislation was part of a yearly state aid act passed in 1991. The act specifies that the test should be prepared by 1993, in three specific subject matters plus others that the state board may specify, and should be based on the state board's model core curriculum outcomes. Students in the graduating class of 1997 must achieve passing scores to receive a high school diploma.

- The director of state assessment had recently resigned and his position had not been filled. The remaining staff, while of high quality, was of limited size, already overworked with existing projects, and had no members with first-hand experience in developing a high school graduation test.

- The legislation did not specify a budget for the required test construction and administration. The state was in a tight financial situation, and no assurance was given that the project would be adequately funded.

Contextual factors will exist in any state, and Michigan's were *not* atypical. It is common to have tight deadlines, legislation demanding that certain subject matters be tested while giving some decision-making power to the state board, tests that must be based on a state core curriculum, and an overworked and underfunded assessment department that is expected to implement the legislation. Thus, the advice given in response to the Michigan legislation should be applicable to any state.

Following this introduction, the report has two major sections. The first section discusses the complex issues that must be faced during implementation. The paper calls attention to a series of issues, then recommends solutions to some of them.

The final section provides an overview of the steps to be considered in developing and implementing a high school graduation test and suggests when these steps need to be taken. Obviously, discussion of the specific procedures to follow, their timing, and the resolution of the issues often overlap.

The Appendix provides a few scenarios for alternative resolutions of various issues. It should be stressed at the outset that there is no single "correct" way to implement a high school graduation test program. Any program will require a series of tradeoffs. Decision-makers must weave their way through conflicting and competing alternatives—each with advantages and disadvantages. Selecting certain alternatives on any given issue will limit the alternatives available for other issues. The scenarios presented in the Appendix are designed to illustrate these tradeoffs.

This report is focuses on the considerable time, effort, and financial support needed for the developmental steps and the initial implementation of the program. However, such support also will be needed in later stages of the program. For example, new items must be written regularly and the entire process requires constant monitoring and evaluation.

Another point to be stressed is that no procedure will produce a perfect assessment instrument or process. Perfect tests simply do not exist. A test should be as good as it can be, given the constraints. No state department of education would necessarily have to follow all of the advice given in this report to produce an acceptable, high quality process/instrument. However, if a state passes legislation requiring a graduation test, it is important that the state department of education, the state board of education, and indeed the state government as a whole maintain a long-term commitment to a high school graduation test and the high quality development of such a test.

Whether any given test or process is legally defensible is ultimately a decision for the courts. If followed, standards established by the measurement profession make a test more defensible. But no set of standards should be used as a checklist.

This report cannot and is not intended to replace the advice that a state department of education will need from an ongoing technical advisory committee. The advice from such a committee is essential to the development of a technically and educationally sound program.

Finally, a common concern is that the resources to develop and implement a program are not present when it must be planned. While the lack of such resources does not preclude the development of a graduation test by a legislative deadline, it certainly makes the task very challenging. It is critical that state departments consider the reasonableness of the plans suggested in this report, given resources (funding and staffing) present in the first fiscal year and likely to be appropriated in subsequent fiscal years.

Issues and Recommendations

Many issues must be considered when implementing a high school graduation test. This section will address several of the more important ones, including core curriculum/test specification, psychometric, educational, legal, policy/administrative, and human/financial resource issues. Many of the issues are connected, and the resolution of one may affect the others.

In preparing this report, we were mindful of legal and professional guidelines that must be considered when designing and implementing a required high school graduation test. Professional standards for tests are articulated in *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985). Much of the legal consideration comes from the case of *Debra P. v. Turlington* (1983, 1984), a broad-based challenge to Florida's high school graduation test requirement.

Core Curriculum/Test Specification Issues

Obviously, one must decide what to test before beginning to construct the test. But the task is not a simple one. Michigan's Public Act 118 of 1991 specified general subject areas, but did not provide sufficient guidelines. Specific decisions need to be made, including the number of questions to take from each sub-area. These decisions are important for educational, psychometric, and legal reasons. This section discusses some of the more important issues and offers recommendations regarding the curriculum to test.

The recommendations (with slight rewording) originated with the Michigan legislation, which reads in part:

> Not later than July 31, 1993, the department shall develop and the state board shall approve assessment instruments to determine pupil proficiency in communication skills, mathematics, science, and *other subject areas specified by the state board*. The assessment instruments shall be based on the state board *model core curriculum outcomes*. Beginning with the graduating class of 1997, a pupil shall not receive a high school diploma unless the pupil achieves passing scores on the assessment instruments developed under this section. [emphasis added] (Subsection 7 of Section 104a of Public Act 118 of 1991)

Specify Subject Matters[2] The state board of education adopted the model core curriculum (Michigan Board of Education, 1991) for a variety of subject matters in October 1991. The model core curriculum student outcomes had been outlined in a two-way matrix, with the subject areas as vertical columns, a cognitive/affective taxonomy of "expectations for

---

[2]Note that all recommendations follow rather than precede the relevant discussion.

students" as horizontal rows, and student outcomes as the cells in the matrix. As one of its earliest decisions regarding test content, the Michigan Board of Education must decide whether it wishes to specify other subject areas for the 1997 requirement. It must recognize that to do so would increase the costs of the assessment *and* make the timelines more difficult to meet. Moreover, it can add and/or change the areas covered by the assessment at a later time. Of course, when such changes are made, students who have already taken the assessments covering the initially selected areas must complete schooling using these "older" versions. When changes are made, newer versions of the assessment must be phased in.

Recommendation 1: The state board should not specify subject areas other than communication skills, mathematics, and science for the initial assessment.

Specify Content Within Subjects After deciding which subject areas to test, one must decide which student expectations and outcomes to assess. High school graduation tests should not sample a state's total core curriculum for measurement, philosophical, and legal reasons. These reasons raise the issue of fairness. Moreover, sampling a total core curriculum would require too much testing time.

One measurement-related reason not to sample the total core is that some of the student outcomes are affective and others are related to team performance, performance measures, or products that, given our current knowledge of assessment techniques, would be difficult and expensive to measure for all students in a fair and reliable fashion. Some content (e.g., speaking/listening skills) would require either videotaping or personally observing all students.

Recommendation 2: Communication skills assessed during the first assessment cycle should be limited to reading and writing.

Test a Sub-portion of the Core Philosophically, not all student outcomes should be assessed, because a core curriculum appropriate for a specific school district is not a necessary domain for *all* students to master. In Michigan, the model core curriculum is not even a requirement for school districts.

Recommendation 3: The state board and the department of education need to determine which subsets of the core curriculum should be included in the assessments. The decision should recognize the importance of students' opportunity to learn the content and some knowledge of what is likely to be in the school curricula by the date of the first test. The total core curriculum is *not* the appropriate domain from which to build the tests.

The legal—and probably most important—reason for not sampling the total core stems from a legal precedent (*Debra P. v. Turlington*, 1984) holding that a student cannot be denied a high school diploma (a property right) *unless it has been adequately demonstrated that the student has had an opportunity to learn the material on the test*. This legal precedent has been incorporated into the professional *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985):

> When a test is used to make decisions about student promotion or graduation, there should be evidence that the test covers only the specific or generalized knowledge, skills, and abilities that students have had the opportunity to learn (p. 53).

Furthermore, in some subject matters (e.g., mathematics), not all teachers are trained to teach all of the material in a state's curriculum. We are not aware of any evidence suggesting that a test required for all students should be based on the total domain of the core curriculum in mathematics.

An essential point to consider is that a core curriculum may be intended to lead the schools, not to reflect what is being taught. This situation is not a problem if school personnel agree that all of the core curriculum outcomes are important and therefore decide to teach them to all students. However, if teachers do not know how to teach certain aspects of the curriculum, legal and ethical problems arise.

Recommendation 4: Once the testable portion of the core curriculum is determined, an administrative rule or statute should specify that the local districts *must* teach this portion of the core.

Determine Opportunity to Learn Although it is typically necessary to delimit the test to a sub-portion of the core curriculum, the test need not and must not merely reflect a "minimal" set of objectives—the "lowest common denominator" of what schools already teach. It is

feasible and defensible to have a test that does "lead" schools to some extent, but the tested objectives must be adequately measurable and students must have an opportunity to learn them. The domain should not be so narrow that one can teach too directly to the test content or format, but it should be defined in sufficient detail so that schools and students know what is expected of them.

Recommendation 5: The testable portion of the core curriculum should be widely publicized in the local school districts. This information should be disseminated in enough detail to make students and educators aware of the knowledge and skills to be tested, without providing so much detail that the students can answer the questions without understanding the curriculum.

Once one determines the portion of the core to be tested, districts and schools must be told to teach this portion. If teachers need assistance in learning to teach the material, the state has some responsibility to help train the teachers. Evidence that the students have had an opportunity to learn the test content must be gathered before the first testing. The contractor could help gather this information during a formal field test.

Recommendation 6: Provide instructional support and training to local teachers if there is a need.

Recommendation 7: Gather evidence from both teachers and students regarding the opportunity to learn the content domain that the tests sample prior to the first administration of the tests.

Phase in Any Changes Once one has determined the testable portion of the core curriculum, developed test items for that portion, and obtained evidence that students have had an opportunity to learn the material, the core and particularly the testable portion of the core should not be subject to frequent changes. When changes are made, it is important not to hold students who are "in the pipeline" responsible for knowing the new content. Changes should affect only students who were below the tenth grade when the changes were made.

Recommendation 8: The state board should not make any changes in the core curriculum or selected testable core prior to the year in which the law first affects graduating seniors.

Recommendation 9: A phase-in period must accompany any changes in the core curriculum, and the tasks described in recommendations 3 through 7 must be repeated.

Psychometric Issues

All participants in the test construction, administration, scoring, and reporting process should be aware of the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985). However, "the acceptability of a test or test application does not rest on the literal satisfaction of every primary standard in this document, and acceptability cannot be determined by using a checklist" (p.2).

This section is divided into subsections on validity, item development, field testing, scoring, standard setting, item sensitivity reviews and bias studies, reliability, scaling/reporting, number of forms, equating, and standardization of administration.

Validity "Validity is the most important consideration in test evaluation . . . [and] refers to the degree to which that evidence supports the inferences that are made from the scores" (AERA, APA, NCME, 1985, p.9.).

Although validity is a unitary concept, evidence of validity may be accumulated in many ways. Traditionally, such evidence has been categorized as content, criterion-related, and construct validity evidence. Construct validity evidence "focuses primarily on the test score as a measure of the psychological characteristic of interest . . . . Such characteristics are referred to as constructs because they are theoretical constructions about the nature of human behavior" (AERA, APA, NCME, 1985, p.9). "Content-related evidence demonstrates the degree to which the sample of items, tasks, or questions on a test are representative of some defined universe or domain of content" (p. 10). "Criterion-related evidence demonstrates that test scores are systematically related to one or more outcome criteria" (p.11). Thus, different inferences that may be drawn from a test score demand different types of validity evidence. It is important not to make insupportable inferences from the scores.

The test name itself may lead to an insupportable inference. For example, if one called it a "functional literacy" test, the name would support the inference that a person who failed the test was illiterate. Thus, the name should be chosen with care so that it does not encourage an insupportable inference about a theoretical construct.

Recommendation 10: The assessment should be named the "[state] high school graduation tests."[3]

Recommendation 11: The department of education should caution its employees and the state board against making any unsubstantiated statements about what the tests measure or what inferences can be made from the test scores. An official statement should be made regarding the tests and the inferences that can be drawn from the scores.

One of the major Standards to be considered is 8.4:

> When a test is to be used to certify the successful completion of a given level of education . . . both the test domain and the instructional domain at the given level of education should be described in sufficient detail, without compromising test security, so that the agreement between the test domain and the content domain can be evaluated (AERA, APA, NCME, 1985, p. 52).

This evaluation should not be left for the test's critics to make after the test has been given. The test developer must work from a content domain that has been sufficiently described and must ensure that the test is an adequate representation of the content domain. This task is typically completed through the judgmental processes of a panel of subject matter experts.

Recommendation 12: Demand that the test developer design sufficient safeguards to ensure that the test adequately samples the defined content.

---

[3]Because different tests will be given for different content areas, we suggest the plural "tests." However, for ease in subsequent writing, we will continue to refer to the total assessment as a test. It should be understood that the reference includes all of the tests.

Besides the question of whether a test matches a content domain, there is always the question of whether the "correct" domain has been assessed. As discussed earlier, opportunity-to-learn evidence helps determine whether the domain is the correct one. In addition, some might argue for construct or criterion-related validity evidence. Because all validity evidence can be considered construct validity evidence, all of the opportunity-to-learn and content validity evidence can be counted as construct validity. In addition, one would want some assurance of such things as appropriate reading levels, lack of item bias, minimal impact of test taking skill, etc.

You may conclude that criterion-related validity evidence is not needed for a high school graduation test. However, a plaintiff could surely find a measurement "expert" who would disagree. This debate stems from the assumption that those who are *able* to graduate from high school are more apt to be employed, to make better employees, or to succeed in college. That is, it is difficult to argue that a high school graduation test has nothing to do with competence in the workplace or ability to function as a member of society. However, the reality is that one can be reasonably uneducated and survive in society.

Test developers and state officials who discuss the test publicly must be careful not to suggest that a test has criterion-related validity for job employment, job success, or college success, unless evidence has been gathered to support such inferences. Such inferences are not necessary to justify graduation standards from an educational institution. For example, no such data have been gathered to support any given required number of credits for graduation or any given grade point average.

Recommendation 13: Be careful not to make any official statements that would suggest that the test has criterion-related validity if supportive data have not been gathered.

Although criterion-related validity is not necessary for a high school graduation test, part of the political motivation for such a test is the assumption that our schools are turning out graduates who are not sufficiently skilled to compete in a global workplace. An intermediate type of validity evidence that could be considered is to obtain judgments (not empirical data) from employers (and perhaps even legislators) holding that the test seems to be assessing areas that they would consider relevant. If such data are gathered, it should be stressed that it is not, technically speaking, criterion-related validity evidence and should not be interpreted as such.

Item Development *If the developed items are faulty, the test is inadequate.* Furthermore, if the original items are faulty, it is extremely difficult to "fix" the test at the field test stage of development. Any item substantially revised following a field test should be subjected to another field test. Thus, it is extremely important to have well-trained item writers. Any

Request for Proposal (RFP) for item/test development must be written to elicit sufficient information from the prospective contractors so that the bid will not be awarded to an incompetent contractor. The Department will need to audit closely the work of the contractor to ensure adequate item development, tryouts, revisions, etc. It is strongly recommended that items be piloted on a small scale before being placed in a large field test.

An issue that may arise is whether in-state teachers should or should not be involved in item writing or editing. Such a policy has both advantages and disadvantages. Using in-state teachers will result in a greater feeling of state "ownership" of the tests, and these teachers may require less training, if the content is somewhat unique to the state. Using in-state teachers increases the chances that test security will be compromised, and may increase the oversight costs if the contractor is located out-of-state.

Another issue is item format. A few critics of multiple-choice items suggest (incorrectly) that such a format cannot tap into higher order thinking skills. While such a charge is not true, it is certainly true that multiple-choice items cannot measure all possible outcomes. Good item writers are able to write appropriate (e.g., tapping objectives beyond factual recall) multiple-choice items for mathematics, science, reading, and, if included, listening. However, areas such as writing and speaking need to be assessed using other formats.

A state department needs to recognize at the outset that it will be expensive to gather performance assessments on every high school graduate. Because there is so much more experience regarding the assessment of writing, and so much more political interest in this area, writing could be a part of a first assessment. Because there is considerably less statewide experience with other performance assessments, initial test development in areas such as speaking is not recommended. Format should not determine content, but cost will determine format—which, in turn, will affect content.

Finally, the state department of education must make a decision regarding how many items to develop initially. While this decision is related to other decisions (such as how many times a year to test, whether any given form can be reused, and whether anchor items are used for equating purposes), two general recommendations can be made.

Recommendation 14: Contract for enough items initially so that after losses through pilot and field testing sufficient items will remain to build forms through the second administration year.

Recommendation 15: Reissue a contract in sufficient time to have items developed and tried out (possibly embedded in a live form) prior to their being needed for the third year.

Field Testing The first live test also can serve as a field test. The more opportunities that the first cohort will have to retake a test, the more acceptable this procedure becomes. The more traditional the subject areas and test formats, and the more experience the test developer has, the more likely this procedure would be acceptable. However, while this can produce an acceptable test, it increases the danger that the test will not be acceptable, and in the worst case scenario it could put the production of an acceptable test behind schedule by a year or more. Thus, a field test prior to the first administration of the test is strongly endorsed. This should be a large-scale field test in which all aspects of the testing process are tested—including the test delivery, administration, security, and scoring processes. (Writing prompts must be field tested.)

Recommendation 16: Schedule a large-scale field tryout for students in the spring, one year before the students who are affected enter tenth grade.

Scoring The scoring of the objective portions of the examination should be contracted to a national scoring service. Commercial contractors have a great deal of experience and are well-equipped to do this scoring accurately and efficiently.

The arguments for and against in-state teacher scoring of writing assessments are much the same as the arguments regarding use of in-state teachers for item writing. Teachers often enjoy and learn from the scoring process. Using in-state teachers to score the papers may either add to or subtract from the credibility of the process, depending in part upon the quality of the training and monitoring process. At any rate, teachers should not be scoring papers from their own or surrounding districts if they could be aware of the identity of the papers being scored.

Strong arguments can be made for using out-of-state personnel to score subjective tests. The major ones are timely scoring and costs. One state has costed out the scoring of writing and found that using classroom teachers is the most expensive option. An "army" of teachers must leave their classrooms for at least four to six weeks two or three times a year. These individuals must be paid their regular rates and substitutes must be provided. More important, however, their expertise is lost during this time. Their students will never have the benefit of that lost instruction. Ways can be found to involve some in-state teachers without the disadvantages—for example, by using teams of teachers to observe the scoring process and using committees of teachers to assist in policy decisions about scoring.

One could, of course, use both in-state and out-of-state scoring and compare the results. While no formal recommendation is offered, the state should carefully consider the alternatives with respect to validity, credibility of results, costs, and ability to receive timely scores.

Standard Setting When using a cut score on a test to determine whether individuals pass or fail, "the cut score becomes the linchpin in the decision process" (AERA, APA, NCME, 1985, p.50). Yet, standard setting is a subjective process, and typically there is dissonance between where policymakers think the cut score should be and the implication of that cut score for the failure rate (i.e., policymakers would typically think the cut score should be reasonably high until they discover that such a cut score produces a "high" failure rate).

Much professional literature exists on the methodology for standard setting. In general, this literature supports the following points: (1) A trained standard setting committee should be involved in making recommendations regarding the standard. (2) This committee should use an iterative process that includes information about the failure rate by major ethnic groups. (3) The impact data should be obtained from the first administration, not the field test. (There may be considerable pressure to set the cut score before the first administration, but the decisionmakers should resist this pressure and hold firm on awaiting live administration results.) (4) The recommendations from the standard setting committee, a description of the process they used, a discussion of the relative costs of false positives and false negatives, and the fact that scores will go up across time should be taken to the group officially responsible for setting the standard, and this group should make the final decision regarding where to set the cut score. The following broad recommendations are made regarding standard setting.

Recommendation 17: Appoint and train a standard-setting committee.

Recommendation 18: Use a technical advisory committee to help develop a specific standard-setting procedure.

Recommendation 19: The state board of education should establish a passing score through administrative rule based upon a recommendation by the superintendent of public instruction with the advice of appropriate committees.

Because the initial failure rate will be greater than the failure rate after the test has been in place for several years, it may be reasonable to set incremental cut scores over time. This allows the cut score to be set so that an inordinate number of students do not fail at the beginning, but the state is not locked into a cut score that is lower than desirable. The advantage of setting these incremental cut scores at the beginning is that it may be easier to do than to reset the cut scores later.

Recommendation 20: Consider setting incremental cut scores for different graduating classes when the state board of education makes its initial decision.

Some additional points about the cut-score process that are mentioned here, but are not made in the form of a recommendation, are the following: (1) Do not put the cut score in the legislation. It is too difficult to change later if such a change is desirable. (2) Report the cut score as a scaled score, not a raw score. This avoids having to explain why different raw scores are set for different forms of the test.

Item Sensitivity Reviews and Empirical Bias Studies All tests should be designed to be free of ethnic, cultural, and gender "bias." There are well-developed methods to eliminate such bias. The first is in the training of the item writers. They should be trained to avoid certain stereotypical words and phrases that may be offensive or may give an unfair advantage to a particular ethnic, cultural, or gender group. A second procedure is to have all items reviewed by a committee of individuals specifically trained to detect items that may show such insensitivity. A third procedure is to compute "differential item functioning" statistics on all of the items based on a field tryout. Those items that are "flagged" by such a statistical analysis should then be brought back to the item bias committee—and probably to the relevant subject matter content committee—for a final determination of whether those items should be removed from the item bank. A fourth procedure is to collect committee members' judgments on whether or not the test as a whole is relatively free of bias.

It is important to note that while the test should be free from "bias," this does not mean that all ethnic, cultural, and gender subgroups should necessarily have the same mean level of performance. If some groups truly have not achieved as many of the skills in one of the subject matter areas (or indeed on a particular item), the test (item) should reflect that true state of affairs. Based on the findings from many previous assessments, state departments of education should anticipate that not all subgroups are achieving at the same level and that the test scores will show those differences. The purpose of the item sensitivity reviews and the differential item functioning studies is to gather data to allow for informed judgments about whether the individual items and/or the test items collectively contain irrelevant content that results in unfairness to a subgroup.

Recommendation 21: The item sensitivity reviews should be completed by a committee that is selected and trained specifically for this task. Most members should represent the state's predominant minority groups. However, it would be wise to include at least one member of the committee who is a minority group member from out-of-state and a recognized expert in this area.

Recommendation 22: Conduct statistical item bias studies. Items that show up as statistically biased should be reviewed (but not necessarily discarded) by an item bias committee (conceivably—but not necessarily—the committee used for the item sensitivity review) and a content review committee.

Reliability Reliability pertains to the amount of test variance that is due to random error. Data should have high reliability, and one whole chapter in the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985) is related to reliability and errors of measurement. While those responsible for monitoring the quality of the test should study these standards, the following specific recommendation is offered:

Recommendation 23: Obtain the following reliability estimates: internal consistency, interrater reliability, generalizability across writing samples, and the reliability or standard error at the cut score.

Scaling/Reporting Once tests have been scored, the students' results must be reported. Generally, it is not considered wise to report the "raw scores" (e.g., number of items right on a test). The scores are typically reported based on some mathematical transformation of the raw scores so that the transformed scores have certain statistical properties (e.g., a specific mean and standard deviation). Because high school graduation tests have not been designed to differentiate among those passing, and because one should not encourage use of information on the differences in students' scores above the cut score (e.g., for employment decisions), one would typically report scores above the cut score only as a "pass."

Other questions arise for those who do not pass. High school graduation tests are typically not designed to be diagnostic, yet many individuals believe that failing students should be given some information that would facilitate efficient and effective remediation efforts.

Thus, the dilemma. Reporting sub-test scores may imply more diagnostic information than can be justified based on such technical considerations as the reliability of the difference scores. However, not to report sub-test scores limits the usefulness of the scores for remediation. Because reporting sub-test scores is a multifaceted and technical issue, it deserves careful attention. If the decision is made to report sub-test scores, it is better to know this fact prior to building the test, because it may have implications for the test specifications.

The issue of which transformed scores (scaled scores) to use for reporting is also a difficult technical issue that cannot be solved in the abstract. Numerous scores could be used. Using the same scaled scores across subject matters does have some advantages. For example, if "200" is designated as passing in Mathematics, 200 could also be designated as passing in other areas. One could equate the cut score, but not the standard deviations (or ranges), or one could equate both. Again, these are technical issues that cannot be resolved in the abstract. Using a common scale across subject areas also may have implications for test development.

Recommendation 24: Scores should be reported as "pass" or "fail." Those individuals who fail should be given some information regarding how close they were to passing, and they should be given some diagnostic information that would facilitate remediation efforts. Important technical details (e.g., reliability of difference scores) regarding various methods of reporting diagnostic information should be worked out and specific plans should be formulated by a technical advisory committee prior to approval of the final test specifications.

Recommendation 25: Consider using a common scale across subject matter areas. This takes some advance planning to avoid adopting a scale that is appropriate for one test but unworkable for another.

Number of Forms

> **Recommendation 26: Develop rules/procedures for designating forms for makeup examinations and out-of-school (i.e., adult education) populations. Determine whether forms will be reused. Determine how many times you will administer the test each year. Determine equating procedures (e.g., number of anchor items). Based on these considerations, develop enough alternate forms to last through the second year of test administration. Develop more forms/ items during this time so that a sufficient supply is continuously available.**

Equating High school graduation test questions need to remain secure and they cannot be reused to any great extent. However, to be fair to individuals who take different forms of the test, the forms need to be equated. It is particularly important that diploma sanction tests be equated *at the cut score*, so that a performance level that was considered a "pass" on one form of the test would not be considered a "fail" on a different form. There are many ways to equate, but the two more common general procedures considered viable for diploma sanction tests are to use anchor items or to pre-equate. Anchor item equating is generally preferable to pre-equating for *final cut score decisions*, because the subareas of the test will likely be differently affected by instructional changes. Pre-equating should be done when initially building various test forms. The cut score will, of course, be set on the original form. The wording of the rule adopting a cut score needs to be carefully considered so that it is clear how to equate that score to scores on subsequent forms of the test.

> **Recommendation 27: Use a technical advisory committee to help develop specific equating procedures.**

Standardization of Test Administration

> **Recommendation 28: Carefully consider policies regarding all test administration conditions. For example, the decision whether or not to use calculators in the math test must be made by the department, not by local school personnel. Train local school personnel adequately to administer the tests. Consider random auditing of the administration process to ensure uniformity throughout the state.**

Education Issues

All of the issues involved in a high school graduation test could be considered educational issues. However, in this section, five special kinds will be discussed: early testing, retesting, remediation, special education, and adult education.

Early Grade Testing If a state is going to have a high school graduation test, it also should conduct tests in earlier grades (e.g., 4 and 8) to assist in identifying students who may not be acquiring prerequisite knowledge and skills at the expected rate. While we support, in principle, state tests in earlier grades, it seems important to call the reader's attention to some concerns. It is surely possible for a student not to have acquired some prerequisite knowledge and skills by, say, grade 8, yet that student—with appropriate effort—may well acquire the knowledge and skills necessary to pass the graduation test. Likewise, passing an 8th grade test that covers prerequisite outcome measures in no way guarantees that a student will acquire the outcome measures sufficient to pass the graduation test. This latter point needs to be made very clear to all students, parents, and educators. The early tests should not and will not cover the outcomes assessed on the graduation test.

> **Recommendation 29: Be cautious about any "predictive" interpretation of the scores of a single individual from testing in earlier grades. Such tests should be thought of as providing only an early awareness.**

Retesting Retest issues are of two types: how and whether to give makeup tests for absentees (not a retest of the same person), and how many chances a single individual should have to pass the test.

If someone is ill or has an excused absence on the day of a test, that person should have an opportunity to make up the test as soon as possible. The state must consider whether the district/building should have a window of opportunity in which it can retain the tests and provide an opportunity for makeup tests. This provision seems appropriate if the window of opportunity is not too long; we suggest approximately one week. Special consideration should be given to the issue of whether alternative forms of the writing prompts need to be used for makeup examinations. Extended absences should be handled on a different basis. Written policies should be formulated regarding all makeup procedures.

Other retake issues include the following: Is the student who fails a test area (e.g., writing) required only to retake the failed area; is a student who fails the test obligated to retake that test during each succeeding administration or may the student "sit out"; and when a school is closed by a crisis (e.g., strike), can the test administration be rescheduled for that particular school outside of the announced "window"?

Recommendation 30: The department should prepare and the board should adopt specific written procedures regarding makeup examination provisions.

The number of permissible retakes also should be a matter of policy. Evidence in other states suggests that four to five total attempts prior to scheduled graduation should be sufficient. A person should be allowed free, unlimited retakes through an adult education program if the person has not passed during the regular high school time period.

Recommendation 31: The department should prepare and the board should adopt specific written rules regarding the number of retakes that should be allowed and how many attempts a student should be given prior to the time that he/she is scheduled to graduate.

Remediation A state that adopts a diploma sanction test requirement should be responsible for assisting the local schools in planning for remediation. It seems wise that a state rule should be established to provide that a child who fails must be given the opportunity for remediation.

Several issues need to be considered regarding remediation. For example, who is responsible for designing remediation materials—the local school or the state? If the state designs the materials, is it responsible for evaluating the materials for their effectiveness? Should the state hold workshops around the state on how to remediate? Should the state attempt to control the publication of materials by commercial publishers? If remediation programs increase the costs to the local districts, will they be reimbursed by the state? How can remediation be completed without the negative side effects of tracking or grouping? If a student who has not passed the graduation test requirements but has passed all other requirements decides to return to school for a 13th year, can that student be counted for state aid? Will local schools be required to document their offers of remediation to those who fail?

Recommendation 32: Develop a detailed proposal that addresses questions regarding remediation efforts and the respective responsibilities of the state, the district, and the student for remediation efforts.

Special Education and Limited English Proficiency The state board of education needs to decide what to do with special education students. This decision includes the possibility of exempting such students and providing special administrative procedures and adapted versions of the test for certain handicapping conditions (see The Americans with Disabilities Act, 1990).

Decisions also must be made regarding how to deal with students who have limited English proficiency (LEP). For example, must they pass the test to graduate; may the tests be administered in the student's first language; may an LEP student be excused from taking the test until he/she has demonstrated some proficiency in English; and should parents' approval be required for such an excuse to be made because schools may want to excuse students to make the school's results look better?

The attorney general's office should be consulted when developing the policies about both special education and limited English proficiency concerns.

> **Recommendation 33: Enact an administrative rule regarding testing issues related to special education students and students with limited english proficiency.**

Adult Education Although any given legislation may not explicitly address the issue of adults' receiving an adult education diploma, we assume that the intention of such legislation would be to include these individuals under the graduation test requirement.

> **Recommendation 34: Individuals in adult education programs who wish to receive high school diplomas should be required to pass the high school graduation test.**

Legal Issues

Any high school graduation test should be built so that it is technically sound. Furthermore, decisions made from the data should be applied fairly. Generally speaking, if one can provide evidence regarding those issues, the process should be legally defensible. Thus, we have already addressed legal issues and will continue to do so in sections following this one. However, some more specific legal issues should be kept in mind and are addressed in this section.

First, the state should be aware that tests are frequently questioned from a technical standpoint. The courts will use the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1985) during any hearing, and it is well to follow those standards as faithfully as possible. With respect to legal issues, it is wise to obtain legal involvement early from the attorney general's office.

Recommendation 35: Obtain the services of the attorney general's office early on in the process and continuously as new policies are developed and implemented.

Recommendation 36: The state superintendent of public instruction and the state board of education should work with the legislature to adopt statutory authority for the high school graduation testing program.

A thorough investigation of liability issues should be made. Do existing state statutes protect employees? If the state department retains the service of local educators, does any state statute protect them? Can a teacher be sued because of a claim that he/she did not teach some content—or teach it well enough? Are committee members who make recommendations covered under state statutes?

Recommendation 37: Carefully investigate liability issues with assistance from the attorney general's office. Attempt to obtain necessary statutes with respect to liability. Inform all committees and all staff regarding their potential liability.

One of the main legal issues other than test quality is due process. Individuals need sufficient notification of the graduation requirement. This notification should be detailed with respect to the domains that tests will cover. Details concerning how to notify students and parents need to be worked out. Certified letters need not be sent to every child/parent. Nevertheless, there should be some documentation that the notices were sent (announced). Procedures such as placing notices in a student handbook, placing notices on report cards, etc., should be considered. One suggestion is to produce a videotape to show all students and have each district provide an affidavit that they have shown the tape to all ninth graders.

Recommendation 38: Schools should be notified immediately regarding this graduation requirement and the information disseminated to all teachers. Students and their parents should be notified no later than the year in which affected students are in the ninth grade.

The general issue of documentation also needs some attention. The lack of various types of documentation can become a central focus of a law suit. For example, when committees review items for sensitivity or bias, a complete record should be kept regarding which individuals considered which items biased and what changes to the items resulted, if they were revised. One also needs to consider how long any documentation should be kept.

Recommendation 39: The department should prepare, and the board should adopt, detailed policies regarding what should be documented and how long the documentation should be kept on file. A general suggestion is that all documentation be kept for a period of at least five years following the school year in which the test was administered. Consider keeping "forever" the initial development documentation and records about when, why, and how procedures are adopted and/or changed.

Rules should be drafted regarding what constitutes inappropriate, unethical, unprofessional, and possibly illegal behavior on the part of educators and students with respect to violating administrative standards, security procedures, and so forth.

Recommendation 40: In consultation with the attorney general's office, and based in part upon discussions with representatives of state education associations (e.g., teachers' unions and administrators' associations), the department should prepare and the state board of education should adopt rules on what constitutes inappropriate behavior on the part of educators or students with respect to test-taking, security issues, and so forth, and what penalties will be imposed for violation of these rules. These rules and the penalties should be disseminated to educators and students prior to the initial administration of the graduation test.

Policy/Administrative Issues

A plethora of policy/administrative decisions must be made and rules must be passed to commence a high school graduation test requirement. These include criteria and approval for committee appointments, test frequency and timing (including early awareness tests), rescoring policies, what to do about transfer students, how to handle RFPs, and security issues. To handle all the requirements of a graduation test adequately requires additional staff time, and a department of education needs to consider carefully whether current staff is sufficient.

The following recommendations are made concerning these issues. Most of these recommendations are either self-explanatory or the rationale for them has been covered in previous sections of this report.

Recommendation 41: The department needs to develop a complete list of rules/regulations that need to be adopted and decide whether they can simply be adopted by the board or whether they need legislative approval.

Recommendation 42: Detailed security arrangements need to be developed.

Recommendation 43: Detailed policies regarding security violations need to be established. Staff should investigate current laws regarding freedom of information exclusions, and, if they are insufficient, request new legislation to exempt secure test materials from the freedom of information regulations.

Recommendation 44: The department needs to determine what additional equipment/facilities are needed for storage of secure materials, shredding out-of-date secure materials, and so on.

Many administrative decisions need to be made and should, once made, be communicated to parents and students. One way to communicate this information is through an annual test administration plan. The plan should follow these general guidelines: the tests should be timed, but with very generous time allotments; the order of administering the tests should be constant across the state within any one administration; a student should take no more than four hours of tests in a day; the tests should be scheduled only during the regular school day; a student should be allowed to take any given test only once during a test administration period; transfer students from other states should be required to take the test unless they have passed a high school graduation test in the state from which they transferred.

The following recommendations are made regarding policy issues. Recommendation 46 suggests administering the test first in the tenth grade. This recommendation assumes that all (or at least most) of the content in the test will have been covered by the end of tenth grade.

Recommendation 45: An annual test administration plan should be developed and disseminated to all school districts.

Recommendation 46: The tests should first be administered to tenth graders in the spring and should be administered at least twice each in the junior and senior years.

Human and Financial Resource Issues

Legislators cannot be expected to recognize the costs of implementing a high school graduation test, and a state department of education must provide a rationale to them to support any request for additional human and financial resources. This section discusses needs in staffing, advisory committees, contractors, and financial resources.

Staffing Needs Typically, the existing department staff available to devote time to an ambitious program of this nature is inadequate. Even with a large proportion of the work contracted to external agencies, there remains a great deal of additional work that must be done by staff. For example, an individual should be assigned major responsibility for each content area to be assessed. A measurement specialist with technical background should spend time writing RFPs. Specific tasks for the contractors need to be developed and the contractors' execution of these tasks need to be monitored. Someone must coordinate the assessment staff in the areas of test development, test administration, and test use and reporting. There needs to be an overall supervisor. The following recommendation is presented with respect to staff needs.

Recommendation 47: The department should conduct a careful study to assess additional staffing needs in assessment and instructional programs.

Advisory Committees The need for several advisory committees has already been discussed, and further information about the composition of these committees can be found in the next section. However, for the ease of individuals interested in human and financial resource needs, they are listed here under a specific recommendation.

Recommendation 48: The following advisory committees should be appointed: a department of education steering committee, a testing policy advisory committee, a bias review panel, a technical advisory committee, a content review committee in each content area of the test, an overall content review committee, and a standard-setting committee.

Contractors Other states have found it essential to employ outside contractors to complete many of the very time-consuming tasks necessary in building, administering, scoring, and reporting the results of a high school graduation test. General experience suggests that there are advantages in not having a large number of separate contractors for separate tasks.

Recommendation 49: Use at most two contractors: one for test development and formal field tryouts and another for test administration, scoring, and reporting.

Financial Resources The need for appropriate staff, advisory committees, and outside contractors relates to financial needs. The specific costs depend on decisions regarding many of the issues already discussed in this report. Costs under some test designs easily can be more than triple what they would be under other designs. Two specific issues that have not been considered earlier and may have cost implications are (1) whether nonpublic school students will be tested and, if so, who will pay the cost, and (2) whether the state is responsible for the financing of state-required local school functions.

Other states can provide detailed information about various costs, and we urge any state considering the development of a high school graduation test to contact several of them. For example, several years ago, one state had an initial cost of $358,000. This amount was for a two-year contract that included the development of an item bank (including extensive field testing) for four tests, four forms each, with a 1/3 anchor overlap across forms. In addition, they allocated $50,000 for a technical review panel and another $50,000 for other committees. Furthermore, they needed one staff person assigned full-time to coordinate each of two contracts, plus other personnel to assist in the program. (Note that no test administration, scoring, or reporting costs are included in the above amounts.)

Another state paid from $35,000 to $50,000 for the development of a set of specifications for each test. Their item development costs were approximately $150 per item. Their test administration costs (assembling, printing, distribution, scanning, and reporting) were about $1.75 per student for their multiple choice tests. The cost of scoring the writing samples using one essay and two raters (with a third judge, if needed) was about $5 per person. This state did not pay its in-state advisory committees (except the in-state member of its technical advisory committee). The staff for implementing the program comprised five individuals, including the supervisor. Although the two states used somewhat different development procedures, the costs came out about the same.

Based on the costs of two other states, we offer the following "ballpark" estimate for costs of test development over the first two fiscal years based on a mathematics item pool of 300 items, a science item pool of 300 items, and a communication (reading and writing) item pool of 400 items. If additional subject matters are to be tested, costs would be greater.

Estimated Costs

Item specification development	$120,000
Item development (1,000 items @ $180)	$180,000
Field tryouts	$150,000
Committee expenses (Honoraria for some, travel for all)	$100,000
Other expenses (e.g., documents, materials)	$100,000
Total (not including staff)	$650,000

NOTE: These costs do not include costs for scoring, reporting, and so on, except for the field tryouts. Nor do they indicate start-up costs, ongoing costs, and costs of other activities that may occur. Part of the costs associated with the administration of the tests must be absorbed in the previous fiscal year, because any administration, scoring, and reporting contract would typically be of such length that it would cross two fiscal years. Therefore, the actual allocation for the assessment for the first two years will exceed the development costs shown above. One five-year budget plan developed by one of the Michigan Expert Panel members estimated the total appropriations needed by fiscal year, without personnel costs, to be as follows:

- FY1 $415,000
- FY2 $587,000
- FY3 $1,397,000
- FY4 $1,344,600
- FY5 $1,973,511

The purpose of the illustrative figures here is not to propose a budget. A budget should be more detailed and may indeed be larger than the numbers indicated here.

Recommendation 50: Obtain more detailed information from other states with similar programs regarding fiscal needs. Make recommendations to the legislature that are sufficient to cover department needs, and make clear to them that the task simply cannot be accomplished without adequate support.

Sequence of Tasks

In designing a program for a high school graduation test, it is useful to have in mind the total set of processes and approximate completion dates for various activities. In abbreviated fashion, the tasks are listed in Table 1 with some suggested timelines (based on the Michigan statutory requirement that the class of 1997 must pass a test to receive a diploma). Obviously, the suggested sequence and timelines are based on certain assumptions about decisions reached (e.g., pilot and field testing of items). Different decisions would result in different steps/timelines.

It is important to note that many process strands actually run concurrently. Furthermore, missing one or more of the targeted deadlines can mean that all other deadlines following that one are missed and that the program cannot be implemented on time. Both the legislators and the board of education need to understand that a lot of work needs to be done, and that it takes sufficient staff and resources (which currently may not be present) to accomplish what the legislation demands.

Below is one possible sequence of activities that could be carried out to develop and implement the assessment program. It represents a sequence that we believe to be a reasonable approach. Detailed suggestions about how to perform those activities are not present in this section. The text and recommendations in the previous section cover many such details. Other possible scenarios are listed in the appendix and would require different tasks and timelines.

TABLE 1: Sample Tasks and Completion Dates
(Assuming requirements are for the 1997 graduating class)

Task 1: Establish appropriate advisory committees. *Establish committees as soon as possible.*

This task involves determining what committees need to be established, determining criteria for selection of the committee members, soliciting and evaluating the nominations, officially appointing and training the committee members, and maintaining the committees over time. We suggest the following committees with the understanding that it might be wise to have some overlap of committee members:

- **Department of Education Steering Committee** (10 members): Representing the offices whose clients are affected by the program
- **Testing Policy Advisory Committee** (10-20 members): Representing the state education community to advise on policy
- **Bias Review Panel** (10-20 members) mostly from a state's minority groups, but with at least one member from out of state who is a recognized expert on bias issues in tests
- **Technical Advisory Committee** (6-8 members) composed of at least one measurement expert from within the state and at least one individual who has been the director of a similar competency testing program in another state
- **Content Review Committees** (20-25 members each) Composed of content experts in each area of the test
- **Overall Content Review Committee** composed of a subset of the members in the Content Review Committees
- **Standard-setting committee** (15-25 members)

Task 2: Determine which subject matter areas and sub-areas to test. This task may require action by the State Board of Education. *Determine subject matter areas as soon as possible, but certainly by May 1992.*

Task 3: Disseminate information about Task 2 to all students who will be impacted, parents, business leaders, and other relevant constituencies. *Complete before schools let out for the summer of 1992.*

Task 4: Complete test specifications for each test area. *Complete by August 1992.*

Task 5: Hire a contractor for development of item specifications, item/test development, and a formal field tryout. *Complete by October 1992.*

Task 6: Have the contractor complete the item specifications, item writing, informal pilot testing, and item editing. *Complete by May 1993.*

Task 7: Perform content committee review and revisions as necessary. *Complete by September 1993.*

Task 8: Produce camera-ready copy for formal field tests. *Complete by February 1994.*

Task 9: Field test items on tenth graders. *Complete by March 1994.*

Task 10: Prepare and disseminate descriptive information and sample test items to assist in preparing teachers, students, and parents. *Complete by March 1994.*

Task 11: Develop and adopt rules governing test administration, scoring, and reporting. *Complete by March 1994.*

Task 12: Select operations contractor for administration, scoring, and reporting. *Complete by March 1994.*

Task 13: Review and revise items as necessary and select items for the first test. *Complete by September 1994.*

Task 14: Conduct regional seminars for school administrators and testing coordinators on the administration, scoring, and reporting procedures. *Complete by October 1994.*

Task 15: Complete production of all necessary materials for first tests and have them ready for distribution. *Complete by February 1995.*

Task 16: Administer first testing to tenth graders. *Complete by March 1995.*

Task 17: Score, analyze results of first administration, and establish passing standards for the first administration. *Complete by May 1995.*

Task 18: Design and implement a plan for releasing test results to the schools and the general public. *Complete by May 1995.*

Task 19: Review and repeat steps above. Plan extended timeline to include at least two administrations per year for 11th and 12th graders. Include time for equating procedures for future test administrations. *This task should be carried out continuously.*

Conclusions

We have discussed a number of issues, offered a number of recommendations, and presented an illustrative list of tasks to be performed with suggested completion dates, for a state mandated high school graduation test. It is clearly possible to develop a well-designed high school graduation test that meets curricular, psychometric, educational, legal, administrative, and resource requirements. However, as this document has undoubtedly made clear, the task is not easy. For the task to be done well, a variety of steps need to be completed. For these steps to be completed, adequate funding needs to be made available. Only with appropriate funding to complete the task well, will the requirement of a high school graduation test be of service to the citizens of a state.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Eeducation. (1985). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Psychological Association.

Americans With Disabilities Act, 42 U.S.C. Section 12101 et seq. (1990).

Debra P. v. Turlington, 564 F. Supp. 177 (M.D. Fla. 1983), aff'd, 730 F.2d 1405 (11th Cir. 1984).

Michigan State Board of Education. (October, 1991). *Model core curriculum*. Lansing MI: Author.

Appendix: Options

This appendix outlines several subject matter and tryout/timeline options. They are provided to assist readers in understanding the inter-relationships between various decisions that may be made. Note that our listing of these options is not to be considered as an endorsement.

Subject Matter Options

1. Test only a sub-core of the core curriculum. This sub-core would contain objectively scored tests in mathematics and science, and a combination of an objectively scored test and a writing sample in communications. [Note that this is the option recommended in the report.]

2. Test the sub-core as listed in (1) above, plus build some assessment instruments for speaking, and some non-objectively scored mathematics and science processes. Have the objectively scored test given under secure conditions as would be done in (1) above, have local teachers also administer the non-objectively scored assessments under secure conditions, and do the scoring of these.

   Advantages: Increases the tested domain and includes content not assessable by multiple-choice items.

   Disadvantages: Increases costs of instrument development and places a burden on the local teachers that they may not wish to carry. There would be a question of credibility regarding the teacher scored material.

3. Build instruments as in point (2) above, but have the scoring done centrally (or at least not by the local district personnel).

   Advantages: Covers the larger domain and provides more credible results.

   Disadvantages: Increased costs of instrument development and scoring.

Tryout/Timing Options

1. Proceed as suggested in the example given in the SEQUENCE OF TASKS section of the report. That is, for a 1997 requirement, have a pilot, a field tryout in the Spring of 1994, and plan on administering the test for the first time to 10th graders in the Spring of 1995. Provide the students with four more options to pass prior to graduation: Fall and Spring of 1995-96, and Fall and Spring of 1996-97.

Advantages: Provides better assurance for high quality items and a high quality assessment in general. Provides sufficient advance warning. Test not given until most (perhaps all) of the curriculum on the test has been covered. Provides sufficient number of opportunities.

Disadvantages: Great time pressures to be ready for the pilot and field tryout portions of the process. If field tryout did not count, the statistics will be different than they will be in the first real administration.

2. Do a pilot study, but make the first actual test in the Spring of 1995 serve as both an actual test and as a field test. (Note that a separate field test is needed for the writing prompts.)

   Advantages: Provides more time for the test development process. Saves money.

   Disadvantages: Critics of the exam would argue that a field test should have taken place. The exam could be of poorer quality. One may find an item is flawed after the fact. This may call into question the credibility of the total exam and may result in the test specifications not being adequately covered.

3. Do both a pilot and a field test, but give the first field test to 10th graders in the Spring of 1995 and the first actual test in the Fall of 1995 with 11th graders. Provide four more opportunities, likely one more in the junior year and three in the senior year.

   Advantages: More time to prepare the field test. All those listed in point one.

   Disadvantages: Less time between first real test and graduation, and the data from the first field test would be on 10th graders while the first real test would be on 11th graders.

Obviously many more options could be illustrated. The decision made regarding subject matter options will obviously have an impact on the viability of the timing options and vice versa.

North Central Regional Educational Laboratory
1900 Spring Road, Suite 300
Oak Brook, IL 60521-1480
(708) 571-4700
Fax (708) 571-4716